The construction system of the brain

Demis Hassabis* **and Eleanor A. Maguire***

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London,
12 Queen Square, London WC1N 3BG, UK

The ability to construct a hypothetical situation in one's imagination prior to it actually occurring may afford greater accuracy in predicting its eventual outcome. The recollection of past experiences is also considered to be a reconstructive process with memories recreated from their component parts. Construction, therefore, plays a critical role in allowing us to plan for the future and remember the past. Conceptually, construction can be broken down into a number of constituent processes although little is known about their neural correlates. Moreover, it has been suggested that some of these processes may be shared by a number of other cognitive functions including spatial navigation and imagination. Recently, novel paradigms have been developed that allow for the isolation and characterization of these underlying processes and their associated neuroanatomy. Here, we selectively review this fast-growing literature and consider some implications for remembering the past and predicting the future.

Keywords: episodic memory; imagination; hippocampus; autobiographical;
functional magnetic resonance imaging; episodic future thinking

1. INTRODUCTION

The recollection of episodic memories is widely accepted to be a reconstructive process as opposed to the simple retrieval of a perfect holistic record (Bartlett 1932; Schacter *et al.* 1998, 2008; Conway & Pleydell-Pearce 2000; Rubin *et al.* 2003; Hassabis & Maguire 2007). Recollection relies on a number of component processes. These include a sense of subjective time (Tulving 2002), connection to the self (Conway & Pleydell-Pearce 2000; Gallagher 2000), narrative structure (Rubin *et al.* 2003), retrieval of relevant semantic information (Wheeler *et al.* 1997, 2000; Gottfried *et al.* 2004), feelings of familiarity (Wagner *et al.* 2005) and rich multimodal re-experiencing of the event (Tulving 2002) in a coherent spatial context (Byrne *et al.* 2007; Hassabis & Maguire 2007; Hassabis *et al.* 2007b; Bird & Burgess 2008). From functional magnetic resonance imaging (fMRI) studies, we also know that a distributed and highly consistent network of brain regions supports memory for past experiences. This comprises dorsal and ventromedial prefrontal cortex (PFC), lateral PFC, the hippocampus, parahippocampal gyrus, lateral temporal cortices, temporoparietal junction, thalamus, retrosplenial cortex (RSC), posterior cingulate cortex (PCC), precuneus and cerebellum (Maguire 2001a; Maguire & Frith 2003; Svoboda *et al.* 2006; Cabeza & St Jacques 2007; Hassabis *et al.* 2007a; figure 1). Despite over a decade of activating this network, however, surprisingly little is understood about the contributions individual brain areas make to the overall recollective experience (Hassabis & Maguire 2007).

*Authors for correspondence (d.hassabis@fil.ion.ucl.ac.uk, e.maguire@fil.ion.ucl.ac.uk).

Taking a different approach, several recent fMRI studies compared recall of autobiographical memories with predicting possible personally relevant future events (known as episodic future thinking (EFT), Atance & O'Neill 2001), and found near complete overlap in the brain networks activated (Okuda *et al.* 2003; Szpunar *et al.* 2007; Botzung *et al.* 2008, but see Addis *et al.* 2007). In terms of characterizing the underlying processes and their mapping to specific brain regions, however, it is clear from these studies that only limited further progress can be made by using EFT as a comparison task because it engages all of the same processes as episodic memory and to a similar degree (Suddendorf & Corballis 1997; Schacter *et al.* 2007). Here, we suggest that in the context of real-world experiences, a productive way to investigate recollection of the past and prediction of the future is, ironically, not to study the past or the future at all. We argue that because the core processes underlying prediction of the future can be co-opted by a range of other atemporal cognitive functions, these processes may be best isolated and understood in the context of paradigms where time is not an explicit factor, such as imagining fictitious experiences. We believe that time does not merit elevation to the level of an independent process with a distinct neural signature. Instead, we view the timestamp of an event (whether future or past) as simply the result of a content or goal difference rather than a change in the fundamental processes involved (see Hassabis & Maguire 2007). In order to test this idea, it has been necessary to develop novel experimental paradigms.

2. USING IMAGINATION

Recently, one important tool in the development of novel tasks has been imagination (Hassabis & Maguire 2007). In many ways, imagining new experiences can be regarded as the purest expression of construction.
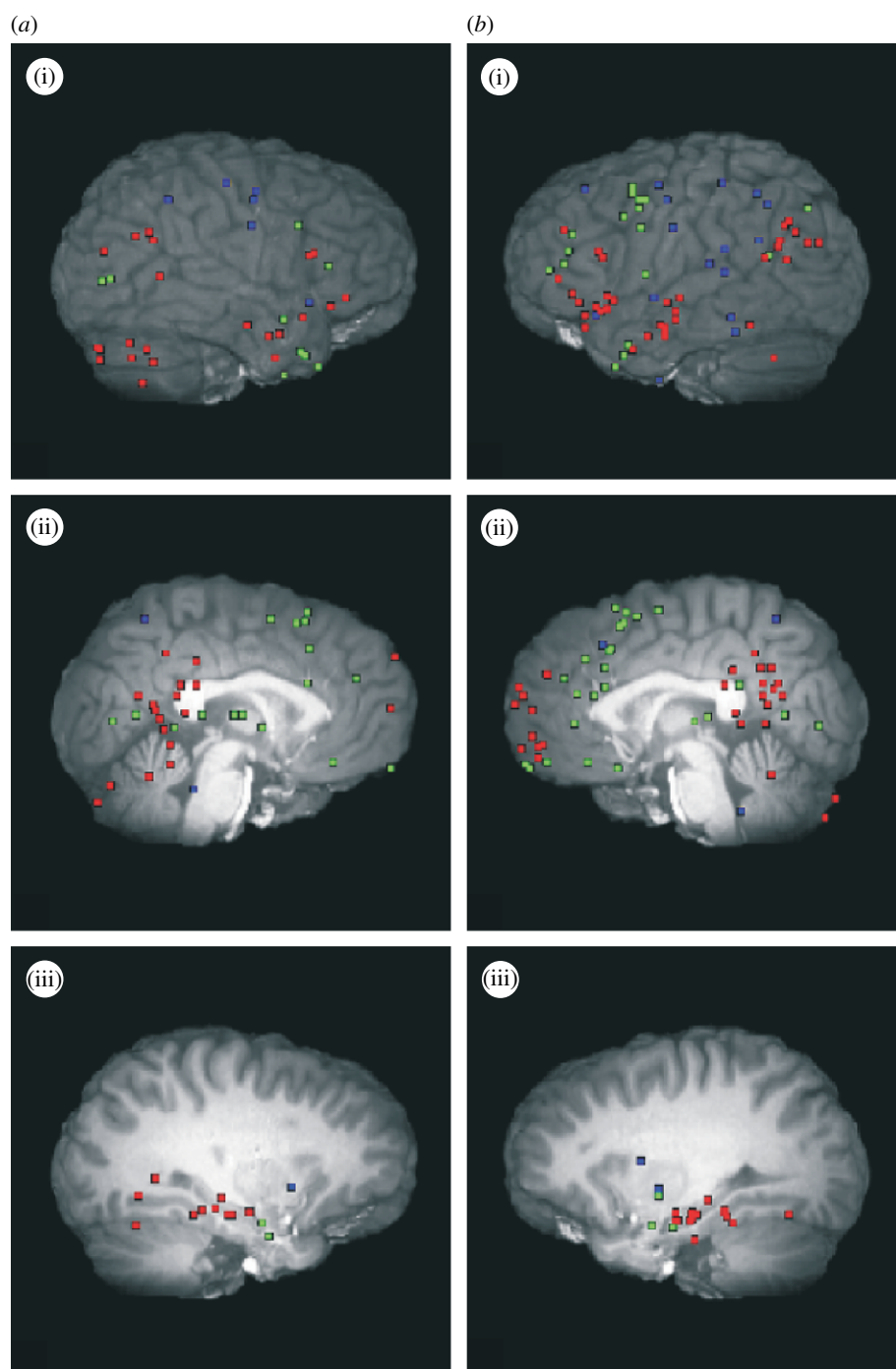
Figure 1. (*a,b*(i)–(iii)) The episodic memory network. Significant peaks of activity from a meta-analysis of 24 neuroimaging studies of autobiographical memory (Svoboda *et al.* 2006). The classic core episodic memory network can be seen in red and includes the hippocampus bilaterally, parahippocampal gyrus, retrosplenial, posterior cingulate and posterior parietal cortices and medial PFC. Activations in core (red), secondary (green) and infrequently reported regions (blue) are depicted across right and left, lateral, medial and subcortical planes. Adapted from Svoboda *et al.* (2006) with permission from Elsevier.

All healthy volunteers can effortlessly use their imagination to a basic degree (indeed humans have told stories and delighted in fiction and narrative for thousands of years), and verbally induced imagination of scenes has been shown to be possible and useful in the neuropsychological context (Bisiach & Luzzatti 1978). Experiences constructed by the imagination, while having much in common with episodic memories, have the advantage of being easier to systematize and experimentally manipulate (Hassabis & Maguire 2007). For example, participants can be asked to construct the same fictitious situations, and their

performances can be compared and contrasted more directly than would be possible in a standard episodic memory recall paradigm (Hassabis *et al.* 2007*b*).

Crucially, tasks involving imagined scenarios can be designed to de-emphasize key features allowing insights to be gained into the neural substrates of these features when compared with episodic memories (Hassabis *et al.* 2007*a*). For example, participants can be asked to construct fictitious experiences in their imagination that are atemporal (i.e. not set in the past or in the future) and with a low connection to the self. Figure 2*a* shows a description of one such imagined

(a)

Cue: Imagine standing by a small stream somewhere deep in a forest

'It's a pine forest. What I can see on the ground all around me are patches of pine needles and brown earth with nothing really growing. The tree trunks are quite narrow. Overhead are the spikes of the green pines and you can only just see the sky.   There's a pine needle smell but down towards the stream there's a slightly rotting smell. It's quite a narrow stream with stones in it and dark water rushing round them causing little white water eddies. There's not much life around the stream and the banks are quite steep sloping down to the stream. It's peaceful and quiet...'

(b)

Time scale: 5 years in future; cue: dress

'My sister will be finishing... her undergraduate education, I imagine some neat place, Ivy League private school... it would be a very nice spring day and my mom and my dad will be there, my dad with the camcorder as usual, and my mom with the camera as usual. My sister will be in the crowd and they'd be calling everyone's name... I can see her having a different hair style by then, maybe instead of straight, very curly with lots of volume. She would be wearing contacts by then and heels of course. And I can see myself sitting in some kind of sundress, like yellow, and under some trees... the reception either before or after and it would be really nice summer food, like salads and fruits, and maybe some sweets, and cold drinks that are chilled but have no ice. And my sister would be sitting off with her friends, you know, talking with them about graduating, and they'd probably get emotional.'

Figure 2. Descriptions of experiences. Representative examples of participant transcripts when cued to describe (a) an imagined fictitious experience (data from Hassabis et al. 2007a) and (b) a personally relevant future experience (data from Addis et al. 2007). Note the absence of explicit temporal and self-relevant statements in (a) that are commonplace in (b) such as 'I will be' and 'my sister is there'.
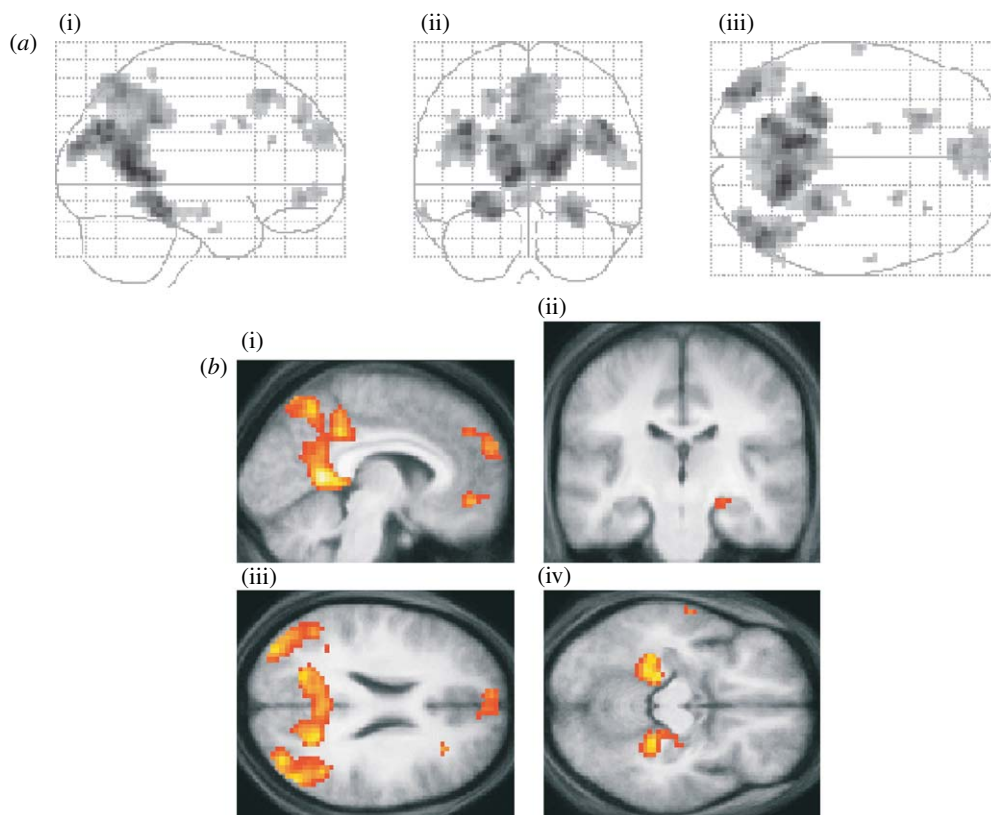

Figure 3. The imagination network. Brain regions active when recalling imagined fictitious experiences that were previously created in a pre-scan interview included the hippocampus, parahippocampal gyrus, retrosplenial and posterior parietal cortices and medial PFC. (a(i)) Sagittal, (ii) coronal and (iii) axial images from a 'glass brain', which enables one to appreciate activations in all locations and levels in the brain simultaneously. (b) Activations on a selection of relevant (i) sagittal, (ii) coronal and ((iii),(iv)) axial sections from the averaged structural MRI scan of the 21 study participants at a threshold of $p < 0.001$ uncorrected (data from Hassabis et al. 2007a).

experience (Hassabis *et al.* 2007*a*). It is interesting to contrast this transcript with one from an fMRI study involving plans for a future personal experience (Addis *et al.* 2007; figure 2*b*). Clearly the imagined scenario is set in the present and in this case does not have the same involvement with the imaginer's self-schema as the personal future event (Conway & Pleydell-Pearce 2000; Gallagher 2000) although both types of scenario involve the adoption of an egocentric viewpoint on the part of the imaginer (Burgess 2006). Being able to manipulate factors such as the level of self-relevance/involvement and the degree of overlap between memories and imagined experiences has the potential to progress our understanding of the core processes and brain areas involved (Hassabis & Maguire 2007).

To this end, we designed a novel imagination task that involved participants richly imagining new fictitious experiences (Hassabis *et al.* 2007*b*). We reasoned that if episodic memory recall was truly a reconstructive process (Bartlett 1932; Schacter *et al.* 1998), with a memory reassembled from its stored constituent components, then some of these integrative processes should also be co-opted by a purely constructive task involving imagination (Hassabis *et al.* 2007*b*). We tested patients with primary damage to the hippocampus bilaterally as this structure is well known to be critical in supporting episodic memory (Scoville & Milner 1957). We found that, as well as being impaired at recalling the past, the patients were not able to richly imagine new experiences. This was the case for EFT scenarios (see also Klein *et al.* 2002; Rosenbaum *et al.* 2005) and, crucially, for constructions that were atemporal and low in self-relevance. Even when all the components necessary to construct a fictitious experience were supplied in the form of visual elements, sounds and smells, patients' performance did not improve (Hassabis *et al.* 2007*b*). The source of their deficit was an inability to integrate the imagined experience into a coherent whole manifesting itself most obviously in the discontinuity of the spatial context. We concluded that the hippocampus plays a critical role in imagination by binding together the disparate elements of an event or scene (O'Keefe & Nadel 1978; Cohen & Eichenbaum 1993).

If the hippocampus plays a critical integrative role in a constructive process such as imagination, it seems plausible that it might also have a similar role in supporting the rich recollection of episodic memories and in predicting the future (Hassabis *et al.* 2007*b*). It has long been known that the hippocampus is required to initially encode the memory of an ongoing event (Scoville & Milner 1957). The traditional view of memory posits that over time these memories are consolidated to neocortex, which is then able to support the recall of remote memories independently from the hippocampus (Squire *et al.* 2004). Conversely, other accounts (Sanders & Warrington 1971; Cipolotti *et al.* 2001; Murray & Bussey 2001; Moscovitch *et al.* 2005; Maguire *et al.* 2006*a*), supported by the results of the majority of fMRI studies on episodic memory (Maguire 2001*a*; Svoboda *et al.* 2006), have suggested that the hippocampus is always required for rich episodic memory recall irrespective of memory age. Various patient studies have been unable to arbitrate

between these two positions largely due to disparate testing protocols, patient aetiologies and scoring systems (Levine *et al.* 2002; Moscovitch *et al.* 2005; Kirwan *et al.* 2008). It has been suggested that discrepancies between studies of remote episodic memory in hippocampal-damaged patients (Bayley *et al.* 2003) might be accounted for by differences in the quality or richness of the recollective experience (Gilboa *et al.* 2004), a feature that is not always captured by existing scoring systems (Kopelman *et al.* 1989; Hassabis *et al.* 2007*b*).

Considering the extant literature above and now also the findings from our imagination study (Hassabis *et al.* 2007*b*), we suggest that the hippocampus may have two distinct functions in episodic memory recall. Furthermore, we propose that such a dual role may help to resolve the long-standing debate about the time scale of hippocampal involvement in episodic memory. First, the hippocampus may be the initial location for the memory index (Marr 1971) which reinstantiates the active set of contextual details (Wheeler *et al.* 2000; Polyn *et al.* 2005; Polyn & Kahana 2008) and later might be consolidated out of the hippocampus (Squire *et al.* 2004). Second, the hippocampus may have another role as an online integrator supporting the binding of these reactivated components into a coherent whole to facilitate the rich recollection of a past episodic memory, regardless of its age. Such a function would be of great use also for predicting the future, imagination and navigation.

Further empirical evidence hinting at a two-process function of the hippocampus comes from structural MRI studies of expert navigators (London taxi drivers) who show increased grey matter volume in posterior hippocampus seemingly at the expense of reduced grey matter volume in anterior hippocampus (Maguire *et al.* 2006*b*). Moreover, their increased spatial knowledge appears to come at a cost to the acquisition of new visual associative information (Maguire *et al.* 2006*b*; Woollett & Maguire 2009). The hippocampus is ideally placed to support these two roles both in terms of the diversity of its multisensory inputs and its specific anatomical properties (Andersen *et al.* 2007), such as the high number of recurrent connections, although clearly more work is required to categorically ascertain if the hippocampus is performing more than one function in episodic memory recall.

3. THE CONSTRUCTION SYSTEM

The (re)constructive process, although critically reliant on the hippocampus, is not supported by it alone. We sought to characterize the entire construction network by using fMRI to compare imagination with episodic memory recall (Hassabis *et al.* 2007*a*). Healthy participants engaged in three tasks while in the scanner: (i) vivid recall of recent real memories, (ii) vivid recall of previously created imaginary experiences, and (iii) construction of new imaginary experiences for the first time in the scanner. Recall of recent autobiographical memories activated the now classic network shown in figure 1 (Maguire 2001*a*; Svoboda *et al.* 2006). Interestingly, imagined experiences were associated with increased activity in many of the same

brain areas (figure 3). A conjunction analysis was performed in order to examine the brain regions activated in common by the three conditions. A distributed brain network was implicated involving the hippocampus, parahippocampal gyrus, RSC, posterior parietal cortices, middle temporal cortices and ventromedial PFC (figure 4). This construction network cannot only account for a large part of the episodic memory recall network (figure 1) and EFT (Addis et al. 2007; Szpunar et al. 2007; Botzung et al. 2008), but also bears a striking resemblance to networks activated by navigation (Maguire 2001b; Burgess et al. 2002), spatial (Maguire et al. 2003; Kumaran & Maguire 2005) and place tasks (Epstein & Kanwisher 1998; Sugiura et al. 2005), as well as those associated with mind wandering (Mason et al. 2007) and the default network (Raichle et al. 2001; Buckner et al. 2008). This suggests there may be a set of key component processes underlying all of these cognitive functions (Buckner & Carroll 2007; Hassabis & Maguire 2007; Spreng et al. in press).

We have suggested that these common processes can be characterized by the concept of scene construction (Hassabis & Maguire 2007). Scene or event construction involves the mental generation and maintenance of a complex and coherent scene or event. This is achieved by the reactivation, retrieval and integration of relevant semantic, contextual and sensory components, stored in their modality specific cortical areas (Wheeler et al. 2000), the product of which has a coherent spatial context (Hassabis et al. 2007b), and can then later be manipulated and visualized. In fact, scene construction is a specific example of 'associative construction', which involves visual imagery, binding and also disparate multimodal elements that, when bound together, (re)create an event as a whole. This includes contextual details such as sounds and smells in addition to visual inputs, people, objects, entities and their actions.

Scene construction differs markedly from 'simple' visual imagery such as that for single objects (Kosslyn et al. 2001), in that it requires the flexible association and integration of many scene elements. Our fMRI study also included tasks requiring vivid visualization of acontextual single objects as a baseline task (Hassabis et al. 2007a). Recalling previously seen or previously imagined objects, or imagining objects for the first time in the scanner resulted in activation of brain areas associated with supporting object representations and manipulations, namely lateral occipital complex and intraparietal sulcus (e.g. Sugiura et al. 2005; figure 5). Moreover, there was no overlap between this simple object network (Sugiura et al. 2005) and that of complex scene construction (Hassabis et al. 2007a), suggesting that they represent dissociable cognitive processes with distinct neural bases. Nevertheless, complex scenes and experiences are clearly constructed out of simpler elements. It has been suggested that past and future experiences draw on information stored in episodic memory (Schacter et al. 2007, 2008). However, we argue that the component elements of constructions are not simply fragments of past events, but can comprise elements that are more abstracted and semantic such as the sound of ocean waves

crashing on the shore or the face of your best friend, and potentially learned over and shared across multiple episodic memories.

Alternatives to scene construction have been proposed (Buckner & Carroll 2007; Schacter & Addis 2007; Schacter et al. 2008). Similar to scene construction, the process of 'self-projection' (Buckner & Carroll 2007) defined as 'the shifting of the self to alternative perspectives in time or space' has been posited as an underlying process common to a set of disparate cognitive functions including episodic memory recall, EFT and navigation. However, while self-projection is clearly an important concept, it conflates several distinct component processes including scene construction (Hassabis & Maguire 2007) and theory of mind (Amodio & Frith 2006). For the purposes of teasing apart the various component processes underpinning episodic memory we suggest it is advantageous to consider constituent processes in as reduced a form as possible. Thus, we believe the construction network is most accurately characterized as being invoked whenever attention is directed away from the current external situation and instead focused inwards towards a rich internal representation of an event, real or imagined. Processes such as theory of mind are only engaged if required, i.e. in the case of EFT or episodic memory recall but not necessarily in imagination or navigation. This may explain why the construction network has a similar pattern of activity to that associated with the default network (Raichle et al. 2001; Buckner et al. 2008) and mind wandering (Mason et al. 2007), cognitive functions that involve minimal external stimuli combined with introspection and rich internal imagery. These constructed scenes or events, created and maintained by the construction network, can then be manipulated further by other processes, such as theory of mind, to allow shifting of the self to alternative perspectives in space or subjective time (Buckner & Carroll 2007; Arzy et al. 2008).

4. ADD-ONS TO THE CONSTRUCTION SYSTEM

We have demonstrated that scene construction is a dissociable set of processes supporting the episodic memory system (Hassabis et al. 2007a), both past and future, but what are some of the other processes that together with the construction system underpin the special properties of episodic memory? We addressed this question using our fMRI imagination paradigm by contrasting the recall of real memories to the recall of previously created imaginary memories matched for difficulty, age, detail and vividness, thus partialling out the effects of the common construction network (Hassabis et al. 2007a). Three distinct areas were more active for real compared to imaginary memories, the anterior medial PFC, PCC and the precuneus (figure 6).

The precuneus has been implicated in studies of recognition memory with increased activity in response to familiar items (Rugg et al. 2002; Wagner et al. 2005; Hornberger et al. 2006; Vincent et al. 2006). Therefore, the increased precuneus activity here probably reflects the relatively greater familiarity of the visualized experience for real memories over more novel
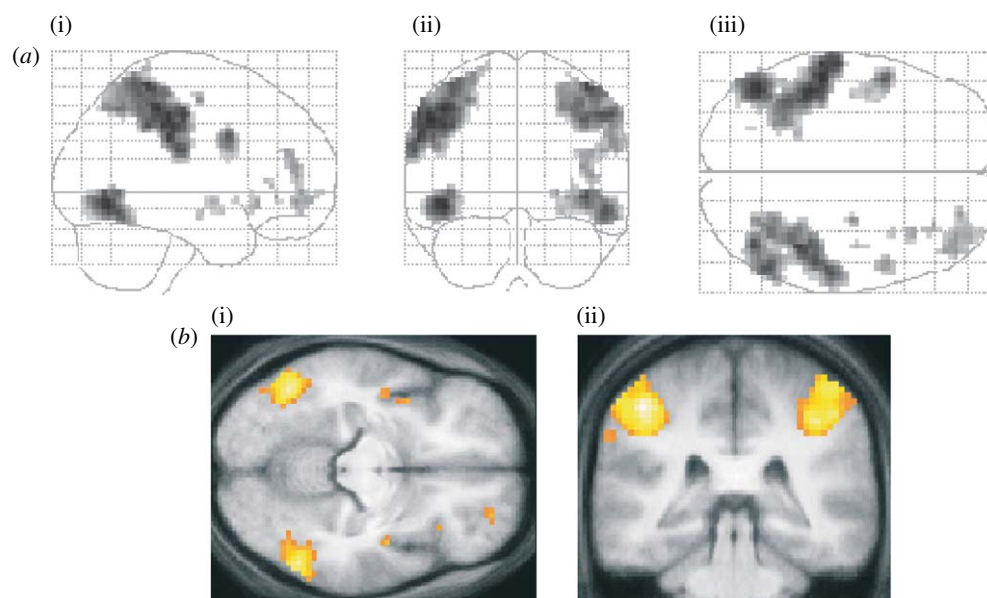
Figure 5. (*a*(i)–(iii)) Neural correlates of recalling single objects. These are the brain areas activated more for recalling (real or imagined) single objects than for complex experiences, and include the lateral occipital complex bilaterally, intraparietal sulcus bilaterally and right lateral PFC. Views of these brain regions are also shown in (*b*) on an (i) axial and (ii) coronal sections from the averaged structural MRI scan of the 21 subjects, at a threshold of $p < 0.001$ uncorrected (data from Hassabis *et al.* 2007*a*).

imaginary memories, given that we controlled for vividness. Activation in anterior medial PFC and PCC is consistent with studies of self-reflection (Johnson *et al.* 2002) and theory of mind (Kumaran & Maguire 2005; Amodio & Frith 2006) suggesting that these two regions support processes related to the self.